JRC Scientific and Technical Reports

Global sensitivity analysis: an approach based on the contribution to the sample mean plot

R. Bolado-Lavin [1], W. Castaings [2], S. Tarantola [2]

[1] Joint Research Centre of the European Commission
Institute for Energy, Nuclear Design Safety Unit
Westerduinweg 3, NL-1755 LE Petten, The Netherlands

[2] Joint Research Centre of the European Commission
Institute for the Protection and Security of the Citizen
Econometrics and Applied Statistics Unit
T.P. 361, 21020 Ispra (VA), Italy

EUR 23433 EN - 2008

JRC EUROPEAN COMMISSION

ipsc Institute for the Protection and Security of the Citizen

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC46545

EUR 23433 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

Abstract

The contribution to the sample mean plot, originally proposed by Sinclair (1993), is revived and further developed as practical tool for global sensitivity analysis. The potentials of this simple and versatile graphical tool are discussed. Beyond the qualitative assessment provided by this approach, a statistical test is proposed for sensitivity analysis. A case study that simulates the transport of radionuclides through the geosphere from an underground disposal vault containing nuclear waste (OECD 1993) is considered as a benchmark. The new approach is tested against a very efficient sensitivity analysis method based on state dependent parameter meta-modelling (Ratto et al. 2007).

Keywords: monte carlo simulation; uncertainty analysis; importance measure; permutation test

1 Introduction

The explicit acknowledgement of uncertainties when trying to understand, predict and control the behaviour of natural and industrial systems is now gaining acceptance and becoming affordable in practice thanks to the tremendous advances in computing capabilities. In the standard probabilistic framework, the uncertain model inputs $X = (X_1, X_2, , X_k)$ and the resulting model outputs $Y = (Y_1, Y_2, , Y_r)$ are treated as random variables characterised by probability distribution functions(Helton 1993). Random or quasi-random sampling strategies are adopted in order to select the model inputs and multiple model evaluations (i.e. Monte Carlo simulation) are used for the propagation of this uncertainty. Subsequently, a detailed analysis of the mapping can be carried out using the input samples and related model realisations.

Sensitivity analysis (SA) is the study of how uncertainty in the output of the model can be apportioned to different sources of uncertainty in the model inputs (Saltelli et al. 2004). Ideally uncertainty and sensitivity analysis should be run in tandem (iterative strategy). Graphical methods are important tools to support, guide and interpret the results provided by sensitivity and uncertainty analysis. While bars, tornado graphs or radar charts can be particularly useful to communicate importance measures, box-and-whisker plots are more suitable for the representation of uncertainty analysis results. Valuable information can also be presented in condensed form by the so-called cobweb plots (Kurowicka and Cooke 2006), which are able to represent graphically multi-dimensional distributions with a two-dimensional plot. Flexible conditioning capabilities facilitate an extensive insight into particular regions of the mapping and a careful analysis of cobweb plots facilitates the characterization of dependence and conditional dependence between inputs and outputs. However, for the visualization of the input-output mapping, the simplest and most widely used plots are the so-called scatterplots. For a given model input X_i and a single-valued model output Y, a scatterplot corresponds to a projection in the (X_i, Y) plane of the sample points defining the (X, Y) hypersurface. Among the possible extensions, model inputs can be plotted against each other with an intensity ramp corresponding to the values of the model response (matrix of scatterplots), and different colours corresponding to different subsets can be used on a single graph (overlaid scatterplots).

Using the classical version of the scatterplot, although a visual inspection can be seen as an empirical and somehow subjective appraisal of pattern randomness, scatterplots provide rich information on mapping, which the other global sensitivity analysis techniques tend to condense into a few sensitivity indices. It is possible to visualize the values taken by the model response Y across the range of X_i. When a pattern can be observed in the scatterplot, the stronger the pattern, the more important the influence of the corresponding input on the model output. Some techniques referred to as grid-based methods can be used to

assess the randomness of the distribution of points across the range divided into bins. Various statistical tests have been developed in order assess common means (CMNs), common distributions or locations (CLs) (Kruskal and Wallis 1952), common medians (CMDs) or statistical independence (SI) (see Kleijnen and Helton (1999a;b), Helton et al. (2006) for recent reviews and comparisons). However, as emphasized by Helton et al. (2006), it is possible that the violation of statistical test assumptions could be leading to misrankings of input importance. In addition, there is no universal rule for the determination of an appropriate division of the range (i.e definition of the grid).

In the Probabilistic System Assessment Group framework, a research group established by the Economic Cooperation and Development (OECD) Nuclear Energy Agency (NEA), Sinclair (1993) investigated changes in the mean and in the variance of various output quantities resulting from finite changes in the inputs probability distribution functions (*ex.* shifts or reduction of their range). An approach was proposed in order to estimate the derivative of the expectation of the analysed model response with respect to the parametrised change of shape. In order to circumvent the difficulties related to discontinuities in the model inputs probability distribution functions, the author suggests to fit a smooth curve to the marginal dependence of the mean of the output on the selected inputs. Although it is not necessary to portray this relation graphically for the adopted approach, the contribution to the sample mean (CSM) plot was recognised as a general tool for sensitivity analysis.

In this paper, the CSM plot, which was not exploited until now according to the authors knowledge, is revived and further developed. In section 3, the scope and potential of this generalised approach are discussed, the outcomes are illustrated using the application example presented in section 2. In section 4, a permutation based statistical test is proposed in order to determine whether the behaviour characterised by the CSM plot significantly departs from randomness. Results from numerical experiments are reported and discussed in section 5, conclusions are drawn in section 6.

2 Description of the test case

In order to illustrate the potential of the plot proposed by Sinclair (1993) and evaluate the reliability of the proposed approach, we consider a model reproducing the behaviour of a radioactive high-level waste repository and the disposed contaminant. The so-called Level E model was used as a benchmark for sensitivity analysis methods (OECD 1993, Saltelli and Tarantola 2002). In this section, the main features of the model will be described and asymptotic Monte Carlo estimates characterising the behaviour of the model will be reported.

2.1 Level E model for radioactive high-level waste repository

The model predicts the radiological dose to humans over geological time scales due to the underground migration of radionuclides from a nuclear waste disposal site. The scenario considered in the model tracks the one-dimensional migration of four radionuclides (^{129}I and the chain $^{237}Np \rightarrow {}^{233}U \rightarrow {}^{229}Th$) through two geosphere layers characterised by different hydro-geological properties. The processes being considered in the model are radioactive decay, dispersion, advection and chemical reaction between the migrating nuclides and the porous medium. The repository is represented as a point source. The simulation model includes twelve uncertain inputs, which are listed in Table 1 together with a set of parameters which are assumed constant.

Notation	Definition	Distribution	Range	Units
T	containment time	uniform	$[100, 1000]$	y
k_I	leach rate for iodine	log-uniform	$[10^{-3}, 10^{-2}]$	mol/y
k_C	leach rate for Np chain nuclides	log-uniform	$[10^{-6}, 10^{-5}]$	mol/y
$V^{(1)}$	water velocity in geosphere's 1st layer	log-uniform	$[10^{-3}, 10^{-1}]$	m/y
$L^{(1)}$	length of geosphere's 1st layer	uniform	$[100, 500]$	m
$R_I^{(1)}$	retention factor for I (1st layer)	uniform	$[1, 5]$	-
$R_C^{(1)}$	factor to compute retention coefficients for Np chain nuclides (1st layer)	uniform	$[3, 30]$	-
$V^{(2)}$	water velocity in geosphere's 2nd layer	log-uniform	$[10^{-2}, 10^{-1}]$	m/y
$L^{(2)}$	length of geosphere's 2nd layer	uniform	$[50, 200]$	m
$R_I^{(2)}$	retention factor for I (2nd layer)	uniform	$[1, 5]$	-
$R_C^{(2)}$	factor to compute retention coefficients for Np chain nuclides (2nd layer)	uniform	$[3, 30]$	-
W	stream flow rate	log-uniform	$[10^5, 10^7]$	m^3/y
C_I^0	initial inventory for ^{129}I	constant	100	mol
C_{Np}^0	initial inventory for ^{237}Np	constant	1000	mol
C_U^0	initial inventory for ^{233}U	constant	100	mol
C_{Th}^0	initial inventory for ^{229}Th	constant	1000	mol
w	water ingestion rate	constant	0.73	m^3/y
β_I	ingestion-dose factor for ^{129}I	constant	56	Sv/mol
β_{Np}	ingestion-dose factor for ^{237}Np	constant	$6.8\ 10^3$	Sv/mol
β_U	ingestion-dose factor for ^{233}U	constant	$5.9\ 10^3$	Sv/mol
β_{Th}	ingestion-dose factor for ^{229}Th	constant	$1.8\ 10^6$	Sv/mol

Table 1: List of model inputs for the Level E.

2.2 Characterisation of the model behaviour

The quantity of interest considered in this study is the annual radiological dose due to the four radionuclides. As emphasized in Saltelli et al. (2004), the dynamics of the total output dose is characterised by two maxima corresponding to the release of two different isotopes ^{129}I (fast dynamics) and ^{237}Np (slow dynamics) respectively. In order guide and corroborate the experiments to be carried out using the CSM plot, some of the results obtained by Saltelli and Tarantola (2002) and Saltelli et al. (2004) will be reported.

Regression and correlation approaches to global sensitivity analysis represent a very simple and intuitive assessment for the (approximate) decomposition of the variance of the variable of interest for linear (or quasi-linear) models. On the other hand, *model-free approaches* like the ones proposed by Cukier et al. (1978) or Sobol' (1993) yield robust and accurate global sensitivity measures (in particular first-order sensitivity indices S_i's) without relying on any assumption on the nature of the mapping between the inputs and the model response. The combined use of the previously mentioned techniques usually provide a valuable insight into the model behaviour. Figure 1 shows the evolution of the obtained coefficient of determination R^2 using regression on the original (for linear effects) and rank transformed values (i.e. R^{*2} for non-linear effects). In order to characterise the importance of interactions, the sum of asymptotic first order sensitivity indices is also provided (interactions are high when the sum of S_i's is small). The temporal evolution of the S_i's for all inputs is described by figure 2.

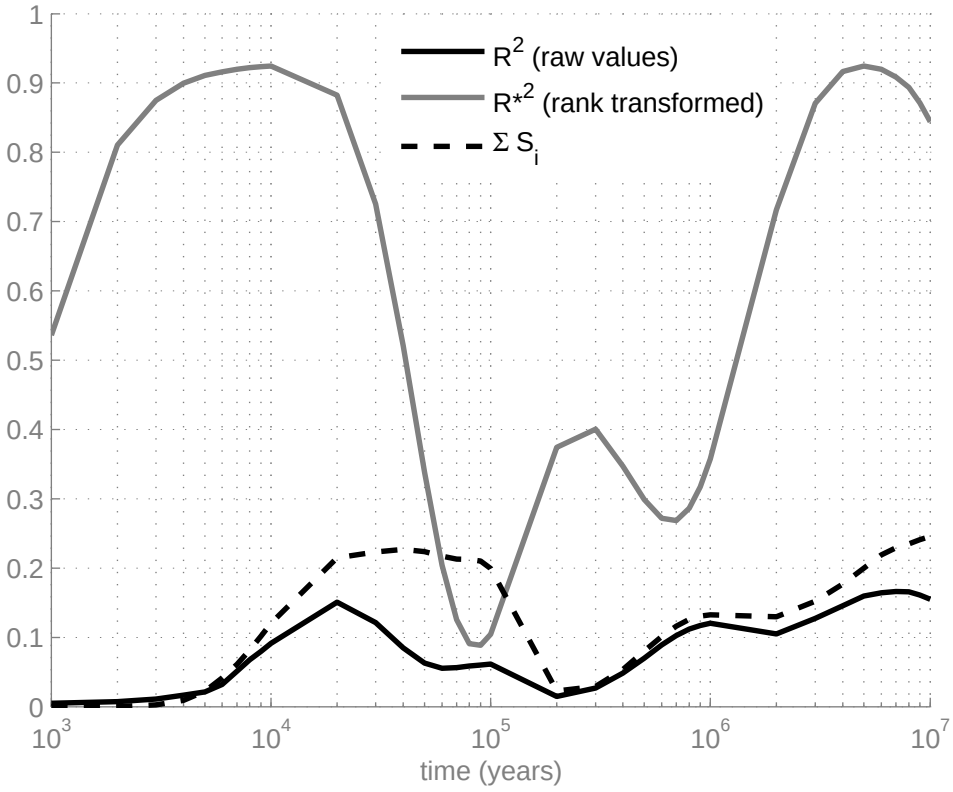

Figure 1: Characterisation of the mapping between the model parameters and the dose at time t using (rank) regression (LHS sample of size 10000) and variance-based sensitivity indices (asymptotic values computed with the Sobol' method)

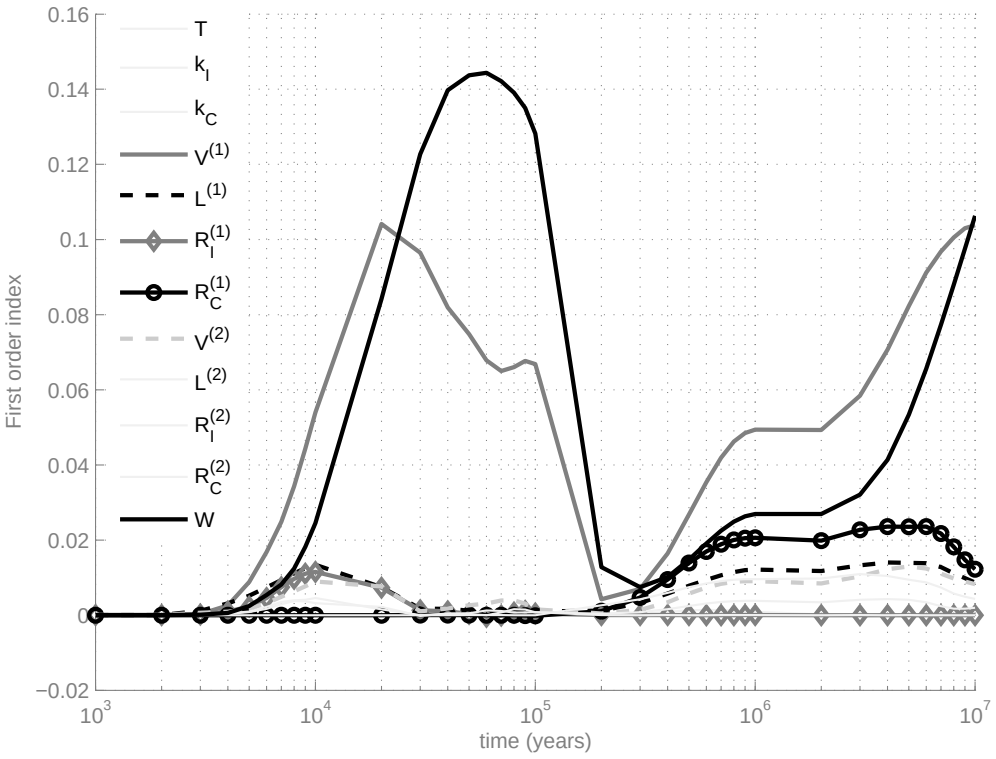

Figure 2: Temporal evolution of first order sensitivity indices (asymptotic values computed with the Sobol' method) for all model inputs (important parameters are highlighted)

The temporal evolution observed for both figures is characterised by two peaks (the second peak being incomplete for the specified simulation period) corresponding to the release of the two different isotopes. The evolution of the first order sensitivity indices for the retention factors of the first geophysical layer corroborates this assumption. The importance of $R_I^{(1)}$, retention factor for ^{129}I in the first geophysical layer is peaking first, then the influence of $R_C^{(1)}$ corresponding to the chain $^{237}Np \rightarrow {}^{233}U \rightarrow {}^{229}Th$ is progressively increasing (see figure 2). Given the difference between R^2 and R^{*^2} (coefficients of determination for the regression on raw values and ranks), the model is highly non-linear for most of the simulation period. Given the sum of first order sensitivity indices (always less than 0.25), interactions also play an important role, especially around 2×10^5 years where both isotopes really contribute to the determination of the total dose. Although the length of the first geophysical layer $L^{(1)}$ and the velocity in the second $V^{(2)}$ play some role, the variability of the total dose is mainly driven by the stream flow rate W and the velocity in the first layer $V^{(1)}$.

3 The contribution to the sample mean plot

In variance-based methods a considerable amount of information on the model output is lost when its mean and variance of a dependent variable is are calculated from a random sample. Considering a general model with k independent inputs $Y = g(X_1, X_2, \cdots, X_k)$, variance-based global sensitivity analysis methods are very efficient in inferring how the variance of the output Y can be quantitatively apportioned to the different model inputs. However, the multidimensional averaging characterizing global sensitivity analysis methods provide only part of the information available from the mapping between the model inputs and the response of interest. In fact, for a given input X_j, it is not possible to assess how a specific quantile of this variable contributes or fails to contribute to the response Y. This is possible using the approach proposed by Sinclair (1993), which is revived and further elaborated in this section.

In order to create a contribution to the sample mean plot, a random (or quasi-random) sample of size N for the inputs and the corresponding model outputs is considered. In order to build the curve corresponding to a given input X_i, the following procedure is adopted:

1. the realisations of X_i are sorted generating the series of values $\{{x_i}^{(1)}, {x_i}^{(2)}, \cdots, {x_i}^{(N)}\}$;
2. the corresponding series of values $\{y^{(i,1)}, y^{(i,2)}, \cdots, y^{(i,N)}\}$ [1] is created ;
3. the ancillary variable M_i is defined, whose sampled values are obtained from the $y^{(i,j)}$, $j = 1, \cdots, N$ as
$$m_i^q = \frac{1}{N} \sum_{j=1}^{q} y^{(i,j)} \quad q = 1, \cdots, N \tag{1}$$
4. M_i is normalised (i.e. $M_i \in [0, 1]$) dividing the values m_i^q by the sample mean of Y;
5. the sampled values of M_i are plotted against $F_{X_i}(x_i)$, the cumulative distribution of X_i (which also lies in the interval $[0, 1]$).

Given the definition of the plot, each point $(F_{X_i}(x_i^{(q)}), m_i^q)$ represents the fraction of the output mean due to any given fraction of smallest values of the input X_i. Therefore, any part of the range on the x-axis corresponds to a quantile range of the selected input X_i. For instance, using the plot , it is possible to assess the contribution to the sample mean of Y of 10 % from the smallest realizations of X_i by analysing the range $[0, 0.1]$ of the x-axis. If the probability distribution function of X_i is uniform, the quantile range corresponds to the same proportion of the range of Xi. More formally, using the approach described previously, estimates

[1] (where $y^{(i,k)}$ is the output obtained when X_i took the value $x_i^{(k)}$, i.e. realisations of Y sorted according to the order of the x_i's)

of the following quantity are represented on the y-axis

$$\frac{\int_{\Omega_{X_{-i}}} dx_{(-i)} \int_{x_i < x_i{}^q} f_X(x) y(x) dx_i}{\int_{\Omega_X} f_X(x) y(x) dx} \tag{2}$$

Equation 2 represents the fraction of the output mean corresponding to values of X_i smaller or equal to its quantile of order q. The subindex $(-i)$ indicates the exclusion of the input X_i. In equation 2, the integrals are computed respectively, over the input space (denominator), over the whole input space excluding input X_i and lastly on X_i up to its quantile of order q. If $F_{X_i}(x_i^{(q)}) \simeq m_i^q \ \forall q$, it means that any quantile range of X_i have a similar influence on the mean of Y. This means that X_i is a non-influential model input, and it can be fixed at any value within its range of uncertainty without affecting $E(Y)$.

In order to illustrate the information provided by such visualization, a Latin Hypercube sample of size 5000 is generated for the Level E model using the inputs pdfs specified in table 1. For four of the twelve uncertain inputs, the marginal dependence of the mean dose is portrayed graphically, for several time points across the simulation period (see figure 3).

For the containment time, whatever the time point considered, all quantile ranges of T have a similar influence on the mean dose. It means that values of T could be assigned at random without having a significant impact on the mean dose. In other terms, the influence of this input factor on the output is not significant. The input T can be considered as a non-important model input. Concerning the stream flow rate, over the simulation period, 60 % of the smallest realizations of W lead to almost 90 % of the mean dose. Therefore, the small values of Y signicantly contribute to $E(Y)$, reducing the upper bound of W would have a very small effect on the mean dose. While the mapping between the input and the dose at time t seems independent from t for T and W, the situation is different for $V^{(1)}$ the water velocity in the first geophysical layer and $R_C^{(1)}$ the retention factor for the chain elements in the same layer. For $V^{(1)}$, while the left region of the range does not significantly contribute to the mean dose for early and late time points, in the middle of the temporal range it is the central part of the range of $V^{(1)}$ which drives variations of Y (inflexion point in the CSM plot). The phenomenon is, once again, probably due to the interactions of both ^{129}I with the chain $^{237}Np \rightarrow {}^{233}U \rightarrow {}^{229}Th$ in determining the total dose. For $R_C^{(1)}$, since the release is much slower for the chain elements, the values of $R_C^{(1)}$ become influent only in the second part of the simulation period (i.e. after 3×10^5 years).

When the input-output mapping is monotonic, the position of the curve in the CSM plot indicates the

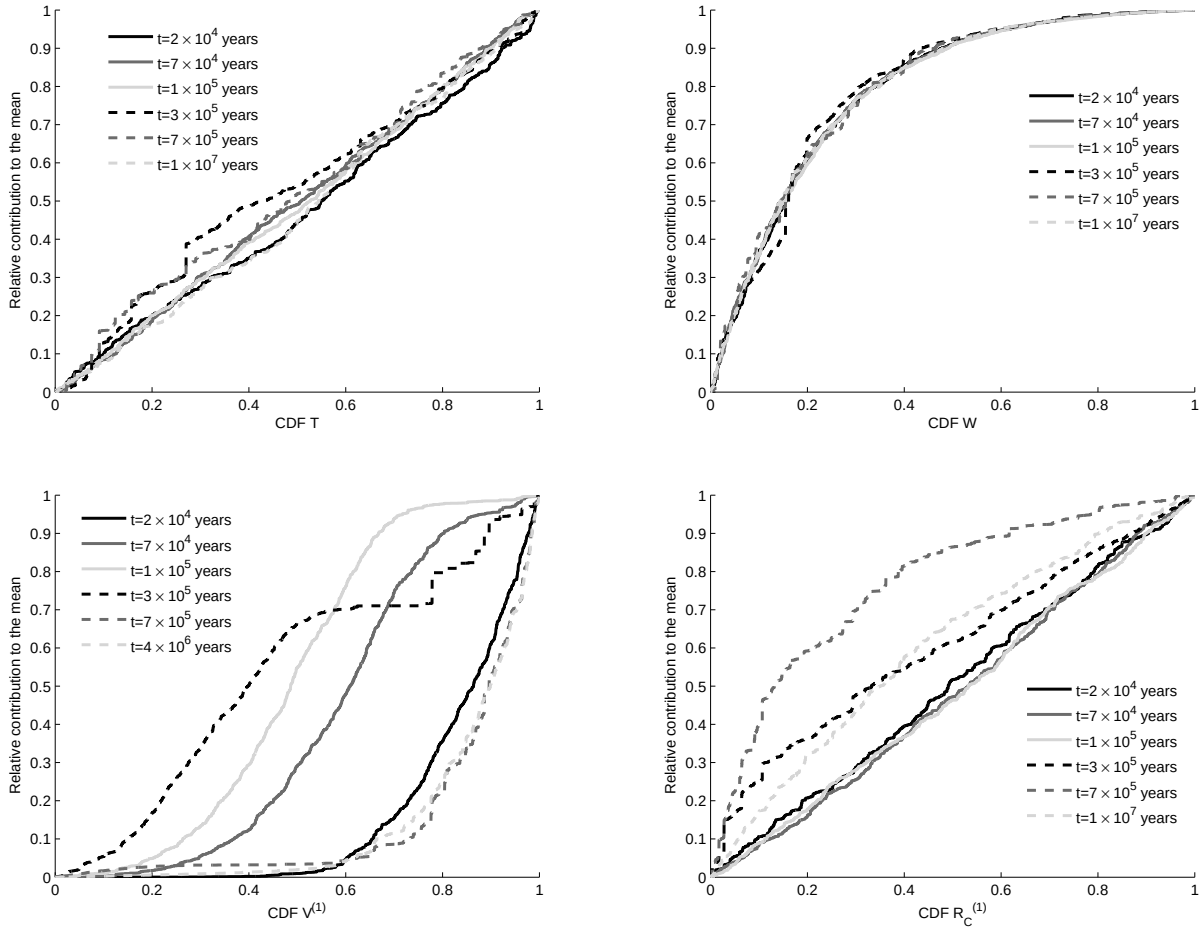

Figure 3: Contribution to the sample mean plot using the dose at different time points for parameters T, W, $V^{(1)}$ and $R_C^{(1)}$

nature of this mapping. A positive monotonic relation will lead to a curve below the diagonal (i.e. small, average and high relative contribution to the mean) and a decreasing relation will be characterised by a plot above the diagonal (i.e. high, average and small relative contribution to the mean). As an illustration, in the model equations (see Saltelli and Tarantola (2002)) the stream flow W is at the denominator of the formula used for the calculation of the dose at all times. The larger W, the smaller the dose at all time points; this is perfectly consistent with the pattern observed in figure 3.

It is very important to emphasise that the steps heights along the $y-$axis depend on the underlying input-output mapping. In fact, the behaviour of the model for a specific response on a particular quantile range for X_i plays an important role. For instance, when the dose at $t = 3 \times 10^5$ years is analysed, the currently used sample size and design do not accurately capture the model behaviour. This happens at the end of the range for the parameter $V^{(1)}$ and at one third of the range for the parameter T. This information is valuable

in order to assess the suitability of the current used sample in order to infer the behavior of the model. As an illustration of the effect of the sample size, the CSM plot for the effect of the parameter $V^{(1)}$ on the dose at time $t = 1 \times 10^5$ years for increasing sample sizes is displayed by figure 4.

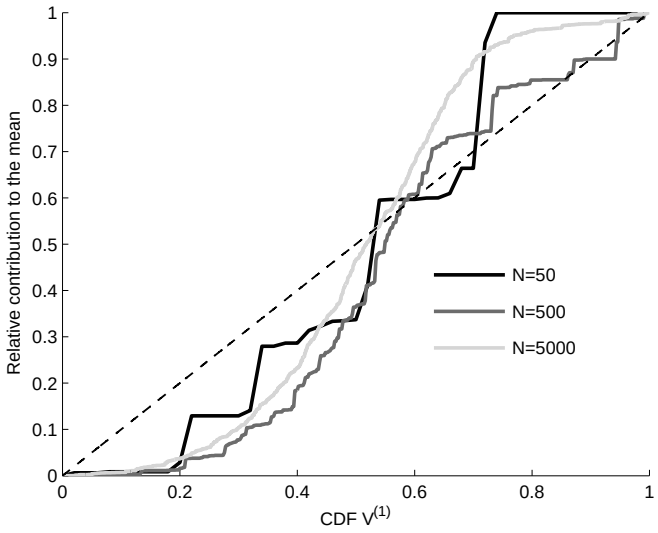

Figure 4: Contribution to the sample mean plot for the parameter V_1 using LHS samples of increasing size

Apart from providing a profitable analysis of the mapping, guiding a possible update of the uncertainty model (i.e. modify the inputs pdfs), support the design of an efficient sampling strategy, assess the distance to convergence for Monte Carlo estimates, the CSM plot can be used for input prioritisation. For the input X_i, the widely used importance measure characterising a first-order effect on the variance of the output is given by

$$S_i = \frac{Var(E[Y|X_i])}{Var(Y)} \tag{3}$$

While the numerator of equation 3 (denoted V_i) characterises the variability of $E[Y|Xi = xi^*]$ across the range of X_i, each curve in the CSM plot characterises the variability of $E[Y|Xi < xi^*]$. There is therefore a correspondence between the numerical importance measures defined by the numerator of equation 3 and the appearance of the CSM plot. If the variance of the conditional expectation is very small, the mean of Y will be very similar across the different quantile ranges of X_i. Therefore, an input featuring a very low first order effect will lead to a line close to the diagonal in the CSM plot. The comparative analysis of figures 2 and 3 corroborates this assertion. Whatever the time point considered, the first order effect of the containment time T is negligible and this translates into curves along the diagonal in the CSM plot. The fact that the

analysis of the CSM plot provides a relative rather than absolute appraisal (V_i rather than S_i) complicates the correspondance for the other model inputs. Important deviations to the diagonal can be observed for W, $V^{(1)}$ and $R_C^{(1)}$ which are significant inputs according to the asymptotic first order effects. Variations in the CSM plot for $R_C^{(1)}$ are consistent with the temporal evolution of $S_{R_C^{(1)}}$ (first order sensitivity index for $R_C^{(1)}$). However, the situation is different for the parameter W. For this input, the CSM plot is very similar for all time points, announcing that V_W might be constant over the simulation period. However, since the unconditional variance is different from one time point to another, this does not translate into a constant S_W (first order sensitivity index for W).

The CSM plot has shown very interesting capabilities for the analysis of any input-output mapping. When the model response does not take only positive values, scale transformations can be applied in order to ensure the reliability of the approach. The technique can be used in order to represent the relation between one input with several outputs (any plot of figure 3), one input with a given model response for several sample sizes (figure 4), several inputs with one output (figure 7). Although the flexibility and versatility of this visualisation technique can be used in various contexts, representing the mapping between several inputs and a given output is really suited for inputs prioritisation. It is precisely the plot which will be exploited in the next section for the development of a statistical test for sensitivity analysis.

4 Development of a statistical test for model inputs prioritisation

In the previous section the CSM plot was used in order to visually and somehow empirically infer the importance of model inputs. The objective of the section paragraph is to describe a statistical test providing a more robust and systematic sensitivity assessment. Rather than trying to detect non-random patterns from the scatter plots (Kleijnen and Helton 1999a), the CSM plot constitutes the primary building block of the proposed statistical test.

The inputs and the output of interest are grouped in a (X, Y) random vector containing $k+1$ components, characterised by its joint multivariate probability density function $f_{X,Y}(x, y)$. An input X_i is completely non-important if the value taken by the output depends only on the values of the other $k-1$ inputs. Under this hypothesis, the conditional distribution of Y given the value of X_i is independent of that value. In other words, this means that the conditional distribution of Y given a value of X_i equals the marginal distribution of Y ($f_{Y|X_i}(y|x_i = x_i^*) = f_Y(y)$), whatever the values of X_i.

When the input X_i is not influential on the output Y, if a permutation is carried out on the realisations

of X_i, since Y only depends on the other (unchanged) $k-1$ inputs, the realisations of Y are not affected by the permutation and the same curve in CSM plot will be obtained. However, when X_i is somehow influent on Y, the permutation will lead to a distinct curve. Let us consider a sample of size N of the vector (X, Y). Computing all possible permutations $(N!)$, all possible CSM curves can be drawn for the pair (X_i, Y).

The rationale carried out in the last two paragraphs characterise the structure of the test to be developed. In order to set up this test, the hypothesis under which the test may be applied is specified, the null and the alternative hypotheses are provided and the test statistic is defined:

- *Assumption*: a sample S of size N for the vector (X, Y) is available. The sample of Y has been obtained via simulation using the sample of X.
- *Hypotheses* (null hypothesis H_0 and alternative hypothesis H_1):
  - H_0: $f_{Y|X_i}(y|x_i = x_i^*) = f_Y(y)\ \forall x_i^* \in R_i$ where R_i is the support of X_i;
  - H_1: $\exists x_i^*, x_i' \in R_i \quad / \quad f_{Y|X_i}(y|x_i = x_i^*) \neq f_{Y|X_i}(y|x_i = x_i')$.
- *Test statistic*: D_m, the maximum vertical distance (absolute value) between the line built according to the procedure described in section 3 and the diagonal. This is the measure of discrepancy with the null hypothesis.

The distribution of the test statistic can be computed using the permutations described previously. However, since the total number of permutations $(N!)$ increases rapidly with the sample size N (ex: $10! = 3628800$), only part of them are carried out in practice. The larger N and the number of permutations considered, the better the approximation of the "maximum distance to the diagonal" distribution. The permutations do not imply any additional model run. It is important to emphasize that since the distribution is calculated from the original sample, a different sample will provide another estimate for the "maximum distance to the diagonal" distribution.

Given a sample of (X, Y), the statistical test for an input X_j can be summarised by the steps described below:

1. Estimate the distribution of the test statistic via Monte Carlo:
   (a) An important number (ex. 10^3) of permutations are carried out for the values of X_j.
   (b) A CSM plot is generated for each permutation.
   (c) The test statistic D_m (maximum distances to the diagonal) is computed for each CSM plot.

(d) The cumulative distribution function of D_m is estimated using standard statistical methods (empirical distribution function and all quantiles via order statistics).

2. Set a critical level α to perform the test (typically 0.05)

3. $D_{m\alpha}$, the value of the test statistic corresponding to α (quantile $1-\alpha$ of the test statistic under the null hypothesis) is computed.

4. The CSM plot is generated with the original sample and the corresponding test statistic D_{mj} is computed.

5. The null hypothesis H_0 is rejected if $D_{mj} > D_{m\alpha}$ (i.e. X_j is an important input)), otherwise it is accepted.

The test statistic is the maximum distance from a line in the CSM plot to the diagonal. Moreover, $D_{m\alpha}$ is the value above which the null hypothesis is rejected. Therefore, the outcomes from the statistical test can be summarized graphically by drawing on the CSM plot a band defined by two parallel lines, separated from the diagonal by a vertical distance $D_{m\alpha}$. In the region within the bands the null-hypothesis is non-rejected. All model inputs characterised by curves in the CSM plot entirely lying in this region are not important.

When the input-output mapping is monotonic, the CSM plot should not include any crossing with the diagonal if an acceptable sample size is prescribed (i.e. small steps heights in the plot). However, when the mapping is not monotonic, the CSM plot might be characterised by one or several crossings with the diagonal. In most practical cases, the situation will be very similar to the behaviour observed in figure 4 where the mapping between $V^{(1)}$ and the dose is clearly non-monotonic. Although the test statistic should still identify $V^{(1)}$ as an important input, the use of the maximum distance to the diagonal might lead to misrankings for important inputs characterized by different types of mappings. A very pathological case would consist in a model leading to periodic mappings for the different inputs. In this case, the maximum distance to the diagonal is not a reliable test statistic. A revision of the definition for the test statistic is proposed in section 5.2.

5 Application of the proposed approach: results and discussion

In order to evaluate the reliability of the statistical test presented in the previous section, the previously described approach is applied to the level E model introduced in section 2. Some numerical results are presented for this specific case and prospects are opened for the treatment of more general input-output mappings.

5.1 Numerical results for monotonic mappings

Given the results obtained in the analysis of the model behaviour (section 2), the response considered here is the dose at 2×10^4 years. For this particular time point, the mapping is monotonic and interactions have an important but not overwhelming influence. In fact, the coefficient of determination of the rank regression is quite high (almost 0.9) and the sum of first-order sensitivity indices is slightly larger than 0.2 (almost the maximum over the simulation period)).

The asymptotic sensitivity indices (see temporal evolution displayed by figure 2) are given by $S_{V^{(1)}} = 0.1042$, $S_W = 0.0842$, $S_{V^{(2)}} = 0.0076$, $S_{L^{(1)}} = 0.0075$, $S_{R_I^{(1)}} = 0.0073$ and very close to zero for the other parameters. Latin Hypercube samples with an increasing number of realizations (N ranging from 50 to 3000) were generated and the statistical test described in section 4 was applied. For comparison, the first-order variance based sensitivity indices are computed using the same samples (i.e. N ranging from 50 to 3000) with the State Dependent Parameter Modelling (SDP) of Ratto et al. (2007), a very efficient method which does not require any specific design for the generation of the input sample. At low sample size, this meta-modelling approach to sensitivity analysis, based on recursive filtering and smoothing estimation (non-parametric smoothing), produces importance measures which are more reliable than those obtained with the Sobol' method (Ratto et al. 2007). Moreover, this technique has shown very good performances in the benchmarking exercise carried out by Gatelli et al. (2008).

The analysis of figures 5 and 6 reveals that at low sample size the ranking provided by the test statistic is more reliable than the one derived from SDP estimates of the first order indices. As far as the ranking of first-order effects is concerned, the main outcome to be identified is that $V^{(1)}$ and W are the most important with a significant advantage for $V^{(1)}$. The sample size needed to obtain this result is four times smaller when using maximum distances to the diagonal derived from the CSM plot (i.e. 500 rather than 2000). However, when the sample size is larger than 1000, the value of the test statistic corresponding to the significance level α (i.e $D_{m\alpha}$) is not longer meaningful. Although the parameters identified as important are correct, when compared to the asymptotic values provided in the previous paragraph, the ranking is not fully consistent and relatively unstable (figure 5). This is not surprising given the asymptotic sensitivity indices reported above.

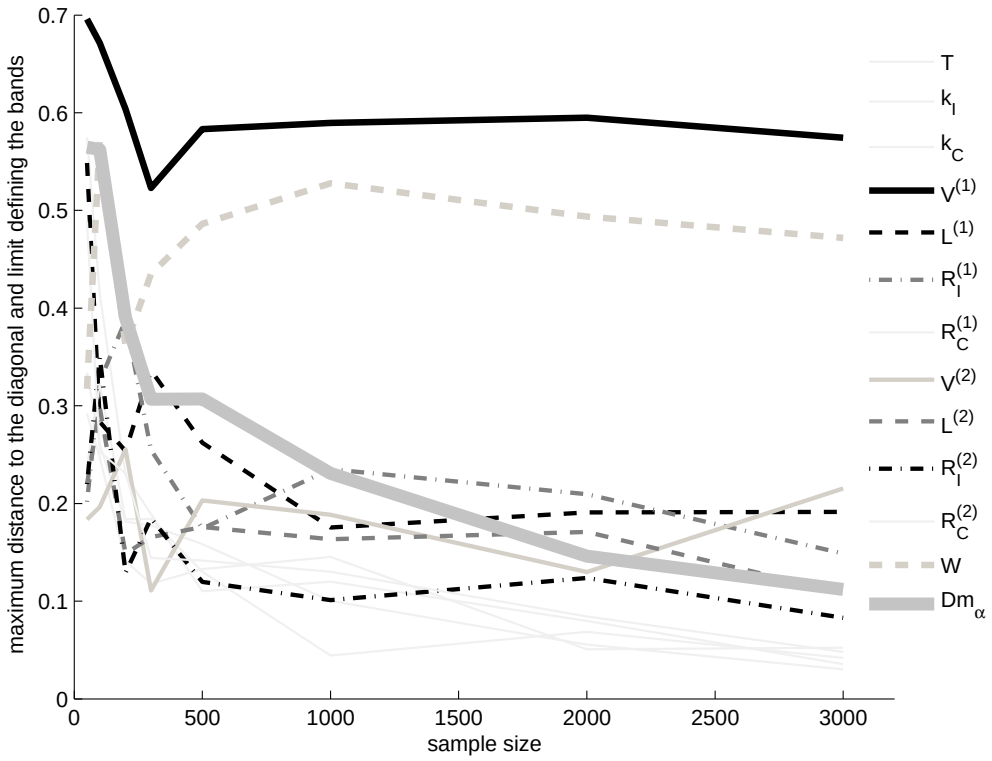

Figure 5: Convergence of the test statistic (maximum distance to the diagonal) and critical value ($D_{m\alpha}$) with LHS samples of increasing sample size (important parameters are highlighted)

In order to asses the stability of the outcomes derived from the CSM plot, 20 Latin Hypercube samples of size $N = 500$ were generated. The variability of the calculated maximum distances to the diagonal across the replicates is provided by figure 8. The fact that the parameters $V^{(1)}$ and W are really influent is a robust outcome. Without requiring any additional model run, the significance of this outcome was already emphasized by the magnitude of the maximum distances observed for those 2 parameters when compared to the test statistic corresponding to the significance level. When the CSM plot lies inside the bands defined by $D_{m\alpha}$ (see figure 7), the corresponding inputs are also characterised by overlapping boxes and/or whiskers in figure 8. However, it is important to underline that using the SDP method, uncertainty bounds are available for the estimates and that the sum of computed first-order effects also informs the user on the remaining part of the variance (i.e. importance of interactions effects).

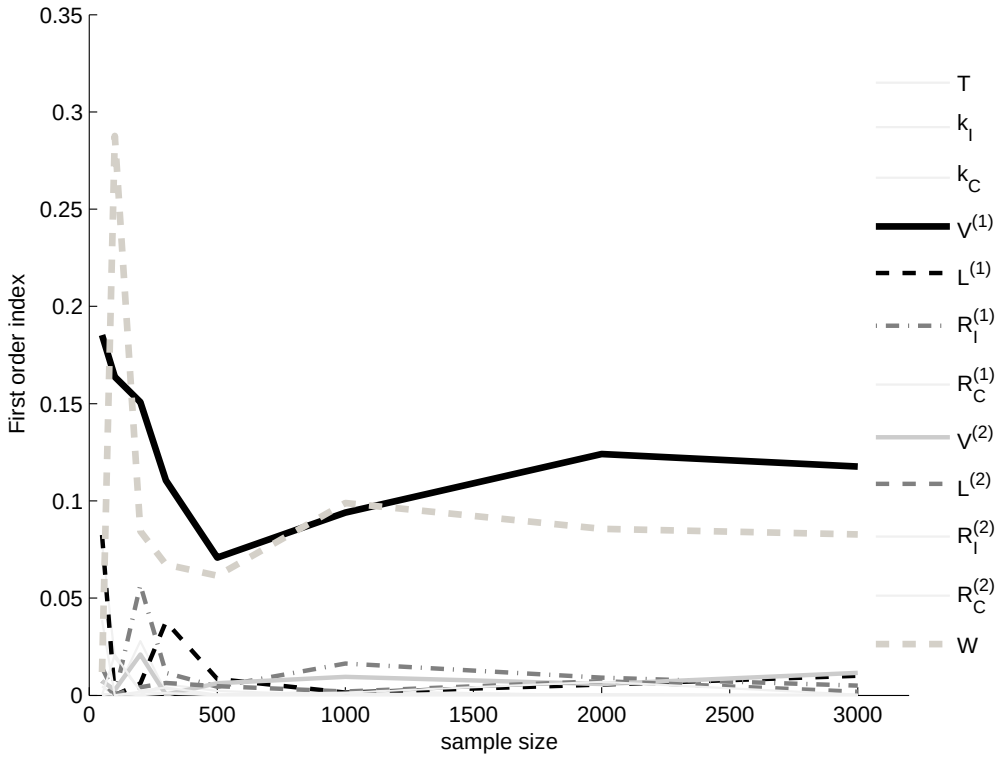

Figure 6: Convergence of first-order sensitivity indices estimated with SDP for LHS samples of increasing sample size (important parameters are highlighted)

5.2 Prospects for a *model free* approach

As mentioned in section 4, when the CSM plot crosses the diagonal, the maximum distance to the diagonal is not a reliable test statistic. Although in practice, a very limited number of crossings will be generally observed, in order to illustrate the current limitations of the approach and open prospects for future investigations a very unfavorable test case (i.e periodic mapping for one of the model inputs) is presented in this section.

The Ishigami function, a non-monotonic analytic function, given by $Y = \sin X_1 + A \sin^2 X_2 + B {X_3}^4 \sin X_1$ with $A = 7$ and $B = 0.1$ and X_i, $i = 1, 2, 3 \sim U(-\pi, \pi)$ was considered. The analytic first order sensitivity indices are given by $S_{X_1} = 0.3139$, $S_{X_2} = 0.4424$ and $S_{X_3} = 0$. The CSM plot, given by figure 9, is characterised by several crossings with the diagonal for the parameter X_2, input factor for which the mapping is periodic. A comparative analysis of figures 7 and 9 confirms that the CSM plot can be used to check the suitability of the currently used sample in order to characterise the model behaviour. While the CSM plot is still characterized by important jumps with an LHS sample of size 500 for the Level E model, the plot is

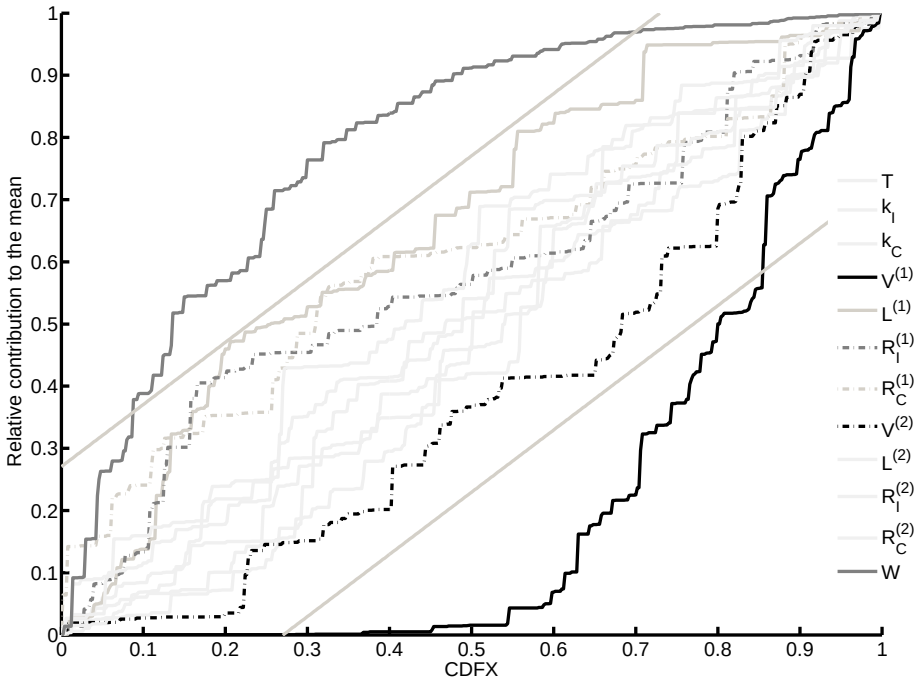

Figure 7: CSM plot for the dose at time $t = 2 \times 10^4$ years using an LHS sample of size 500, the bands define the acceptance region of the null hypothesis (important parameters are highlighted)

already quite smooth for the Ishigami function using a quasi-random sample (Sobol' 1976) of size 128.

The input for which the maximum distance to the diagonal is greater is obviously X_1 and the plot almost coincides with the diagonal for X_3. When the classical version (i.e. with the maximum distance to the diagonal as a test statistic) of the test is applied, the test statistic is given by $D_1 = 0.1135$, $D_2 = 0.0369$ and $D_3 = 0.0143$. The corresponding p-values are $p_1 = 0$, $p_2 = 0.2430$ and $p_3 = 0.9950$. In order to handle more general situations like the one presented here, a natural extension of the test is to use the sum of maximum distances (rather than the maximum distance) for the test statistic. Using this approach, since the CSM plot crosses the diagonal 3 times for X_2, the other distances are unchanged but $D_2 = 0.1372$. The corresponding p-values are $p_1 = 0.0070$, $p_2 = 0.0010$ and $p_3 = 1.0000$. Using the proposed extension for the statistical test, the ranking among the inputs is consistent with the analytical first order indices. However, although the approach has been validated for other analytical models, some numerical problems still remain for a systematic application. They are due to the fact that given the steps potentially characterising the CSM plots at low sample size (see figure 4), it is not straightforward to distinguish the *real modes* from this *noise*. Some investigations are still in progress and will be reported in due course. It is important to underline that using the sample sample, the SDP approach also performs very well for this specific test case.

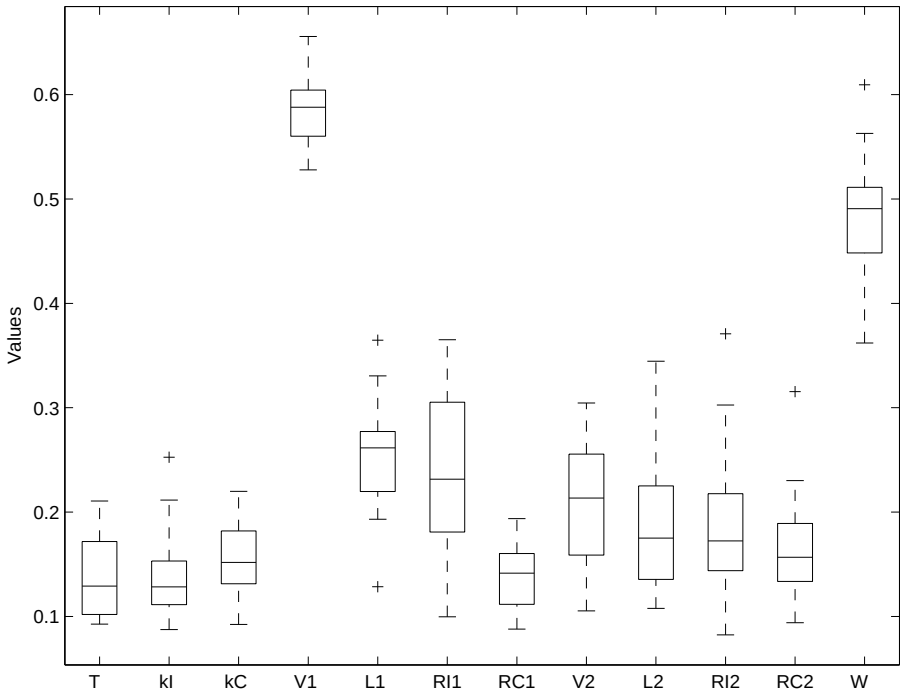

Figure 8: Box-and-whisker diagram showing the stability of the maximum distances to the diagonal using 20 LHS replicates of size 500

The obtained first order indices are $S_1 = 0.3275$, $S_2 = 0.4008$ and $S_3 = 0.0000$, the standard error estimates are respectively 0.0277, 0.0546 and 0.0016. As emphasized in Gatelli et al. (2008), the efficiency of SDP can be optimally exploited with Sobol' quasi-random sequences.

6 Conclusions

The contribution to the sample mean plot has shown an interesting potential for the analysis of the relation between the uncertain model inputs and the resulting model response. The visualisation enables the analysis of the evolution of the contribution to the mean across the range, simultaneously for all model inputs. Therefore, a single plot provides a valuable analysis of the input-output mapping. Moreover, the smoother the CSM plot, the closer Monte Carlo estimates (related to the mean or central dispersion of the output) should be to the asymptotic values. This graphical tool could provide guidelines to improve the sample design or even compose the building block of a variance reduction strategy. Important steps in the CSM plot indicate where additional sample points should be thrown.

For the prioritisation of model inputs, global importance measures can be derived from the CSM plot

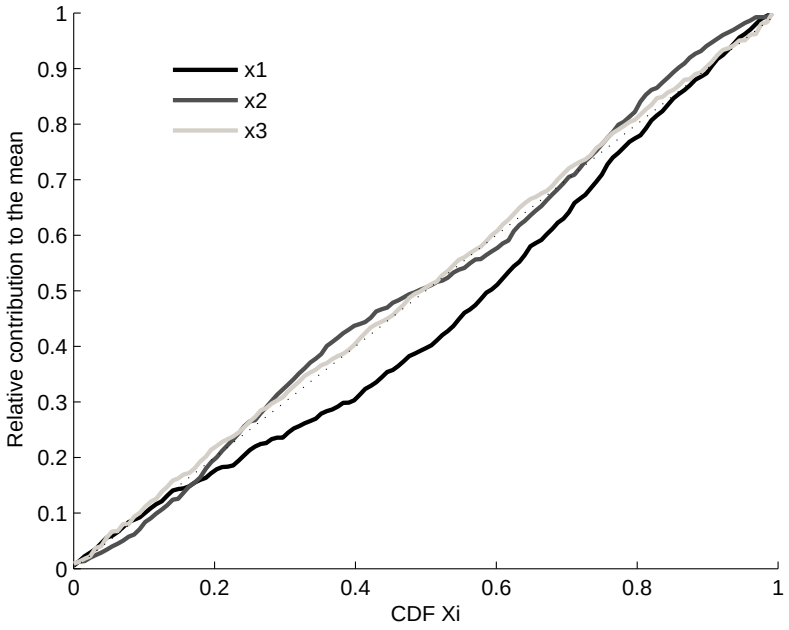

Figure 9: Contribution to the sample mean plot for the Ishigami function using an LPτ sample of size 128

and provide the same ranking like the widely accepted first-order variance based sensitivity indices. Although the CSM plot does not provide variance based sensitivity indices, the significance of the ranking is assessed using a permutation test which does not require any additional model runs. In practice, only a small fraction of the total number of possible permutations can be performed. As long as this amount leads a reliable description of the cumulative probability distribution for the maximum distances to the diagonal, the number of permutations does not have a significant influence on the outcomes. The approach is not prone to type-II error (treating important inputs as non-influential) but might be exposed to type-I error (non-influential inputs as important). Apart from the numerical problems to be solved (see discussion in section 5.2) for some non-monotonic mappings (leading to crossings in the CSM-P), the main limitation of the approach lies in the fact that inputs are ranked with respect to the first order effects but no information is available concerning the remaining part of the variance. On the contrary with summing up first-order effects, in variance based techniques, the analyst can also assess the importance of interaction effects. For the characterisation of second order interactions, an extension of the methodology could be developed using the equivalent of the diagonal for a 3 dimensional surface (i.e. a plane).

In summary, the graphical tool can be used for numerous purposes including the assessment of the direction of change when modifying the inputs probability distribution functions. Within a more classical sensitivity analysis framework, since no particular sampling design is required, the CSM plot and the proposed statistical test can be used in combination with other sensitivity analysis methods for inputs prioritisation. It can

be really reliable and efficient at low sample size if the inputs importance follow a Pareto law (few dominant inputs) but should not be used for fixing non-influential model inputs. Since the construction procedure is straightforward, exploiting the information which could be derived from the contribution the sample variance plot might also lead to interesting outcomes.

Acknowledgements

The authors of this paper are grateful to Marco Ratto from the Joint Research Centre of the European Commission for providing the SS-ANOVA-R MATLAB toolbox (http://eemc.jrc.ec.europa.eu/softwareSS-ANOVA-R.htm) for the global sensitivity analysis based on State Dependent Parameter metamodelling.

References

R.I. Cukier, H.B. Levine, and K.E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26:1–42, 1978.

D. Gatelli, S. Kucherenko, M. Ratto, and S. Tarantola. Calculating first order sensitivity measures: A benchmark of some recent methodologies. *Reliability Engineering & System Safety*, In Press, Accepted Manuscript:–, 2008.

J.C. Helton, J.D. Johnson, C.J. Sallaberry, and C.B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10-11):1175–1209, 2006. doi: doi:10.1016/j.ress.2005.11.017.

Jon C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42(2-3):327–367, 1993.

J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison of techniques. *Reliability Engineering & System Safety*, 65 (2):147–185, 1999a.

J. P. C. Kleijnen and J. C. Helton. Statistical analyses of scatterplots to identify important factors in large-scale simulations, 2: robustness of techniques. *Reliability Engineering & System Safety*, 65(2):187–197, August 1999b.

William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47 (260):583621, 1952.

Dorota Kurowicka and Roger M. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley and Sons, 2006.

OECD. OECD/NEA PSAG User group, PSACOIN Level S Intercomparison. An International Code Intercomparison Exercise on a Hypothetical Safety Assessment Case Study for Radioactive Waste Disposal Systems. Technical report, OECD - NEA publication, 1993.

Marco Ratto, Andrea Pagano, and Peter Young. State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177(11):863–876, 2007.

A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of American Statistical Association*, 97:702–709, 2002.

A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley and Sons, 2004.

J.E Sinclair. Response to the psacoin level s exercise. In *PSACOIN Level S intercomparison*. Nuclear Energy Agency, Organisation for Economic co-Operation and development, 1993.

I.M. Sobol'. Uniformly distributed sequences with additional uniformity properties. *USSR Comput. MAth. Math. Phys.*, 16(5):236–242, 1976.

I.M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993. English translation of Russian original paper.

The Institute for the Protection and Security of the Citizen provides research-based, systems-oriented support to EU policies so as to protect the citizen against economic and technological risk. The Institute maintains and develops its expertise and networks in information, communication, space and engineering technologies in support of its mission. The strong cross-fertilisation between its nuclear and non-nuclear activities strengthens the expertise it can bring to the benefit of customers in both domains.

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Stefano Tarantola
Address: Via E. Fermi, 2749
E-mail: stefano.tarantola@jrc.it
Tel.: +39 0332 789928
Fax: +39 0332 785733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 46545

EUR 23433 EN
ISSN 1018-5593

Luxembourg: Office for Official Publications of the European Communities

Printed in Italy

European Commission

EUR 23433 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen
Title: Global Sensitivity Analysis: an approach based on the contribution to the sample mean plot.
Author(s): Bolado Lavin Ricardo, Castaings William, Tarantola Stefano
Luxembourg: Office for Official Publications of the European Communities
2008 –pp. 21. –
EUR – Scientific and Technical Research series – ISSN 1018-5593

Abstract
The contribution to the sample mean plot, originally proposed by Sinclair (1993), is revived and further developed as practical tool for global sensitivity analysis. The potentials of this simple and versatile graphical tool are discussed. Beyond the qualitative assessment provided by this approach, a statistical test is proposed for sensitivity analysis. A case study that simulates the transport of radionuclides through the geosphere from an underground disposal vault containing nuclear waste (OECD 1993) is considered as a benchmark. The new approach is tested against a very efficient sensitivity analysis method based on state dependent parameter meta-modelling (Ratto et al. 2007).

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.